Concordia University

Assignment 3:

GPT-4o, the Worst Mathematician Ever

Short Essay

Louis Barbier (40254233)

CART 498 - GENERATIVE ARTIFICIAL INTELLIGENCE

Gabriel Vigliensoni

February 5th, 2025

Why Is My Bot Failing

GPT stands for General Purpose Transformer, reflecting its design as a versatile model for various language tasks. However, transformer architectures have a fixed context window, meaning they can only "remember" a certain number of tokens at a time. Complex iterative math often requires maintaining a precise, cumulative state over many steps. Without a dedicated mechanism to preserve and update numerical state accurately, the model can lose track of intermediate values, leading to mistakes.

Additionally, LLM's lack of built-in mechanisms to perform arithmetic operations reliably, they struggle with tasks that require the kind of systematic, error-free computation found in traditional algorithms. They don't "understand" mathematical principles in the way a symbolic math engine does; instead, they rely on learned statistical correlations: each token is still predicted based on probability rather than calculated exactly.

As the number of iterations (i) increases, the model must retain an ever-growing series of intermediate multiplication results; thus, even a single error can quickly propagate, leading to an entirely incorrect final answer. In contrast, raising the base number (n) has a relatively minor impact on the calculator's accuracy, because the fundamental arithmetic operations remain consistent regardless of the magnitude of the base.